Dear Reviewer,

We sincerely thank you for your careful reading of our manuscript and for the constructive comments and suggestions. We greatly appreciate the time and effort you have devoted to reviewing our work. Your insightful comments have helped us improve the clarity and quality of the manuscript.

In accordance with all comments, we have carefully revised the manuscript and provided a detailed, point-by-point response. All modifications in the revised version are clearly marked. The associate editor and reviewers' comments are presented in **black**, our responses in **blue**, and the corresponding revised texts quoted from the manuscript are highlighted in **red**. In addition, we have conducted several additional analyses and experiments to address the reviewer's concerns and further validate the proposed approach. Some of these results have been incorporated into the revised manuscript, while additional results are provided in the supplementary materials for reference.

We hope that the revisions adequately address the reviewer's comments and improve the manuscript. We thank you again for your valuable suggestions and consideration.

Thank you for your valuable comments and suggestions.

Response to Reviewer 1

The size and the extent of the proposed database are remarkable and certainly of interest for the community. However, there are a few issues that must be addressed before publication:

#Comment 1#: extracting 1D profiles from the same 3D geology, while adding some random fluctuation, seems to create a bias in the dataset (profiles are close to each other and they all described the same large geological structures).

#Response 1#: Thank you for raising this important concern. During the dataset construction, we carefully considered the potential bias that could arise when extracting multiple 1-D velocity profiles from the same geological model. To address this issue, we implemented several strategies to maintain structural diversity while preserving geological realism, including: (1) using regional-scale geological models as structural templates, (2) introducing controlled perturbations to expand the model space, and (3) performing statistical analyses to verify the diversity of the generated profiles.

For the OpenSWI-deep dataset, 1-D velocity profiles are extracted from regional-scale 3-D geological models to preserve realistic large-scale geological structures. Although the number of publicly available high-quality 3-D models is limited, these models contain well-constrained geological features such as sedimentary layering, crustal architecture, and regional velocity gradients. Using them as structural templates ensures that the generated velocity models remain geologically meaningful rather than completely random. To mitigate potential similarity among profiles derived from the same 3-D model, we introduce controlled random perturbations to both layer velocities and layer thicknesses. These perturbations are designed to mimic small-scale heterogeneity commonly present in real Earth structures but often unresolved in regional-scale geological models. This procedure effectively expands the model space while preserving the large-scale structural characteristics inherited from the source geological models.

To further evaluate the structural diversity of the generated models, we performed additional statistical analyses. Specifically, we randomly sampled 100,000 pairs of velocity profiles and computed the distribution of their L2 distances. In addition, Principal Component Analysis (PCA) was applied to visualize the structural distribution of velocity models derived from different source datasets. The results (Fig. D1) show that the generated profiles span a broad range of structural variations. Profiles derived from different regional models occupy distinct regions in the PCA space, while profiles originating from the same model still exhibit substantial

variability due to the perturbation strategy. These results indicate that the dataset does not collapse into a narrow cluster of highly similar profiles but instead covers a wide spectrum of plausible velocity structures.

For the OpenSWI-shallow dataset, potential bias is further reduced through a controlled profile sampling strategy from 2-D geological models. Different sampling densities are adopted according to the structural complexity of each geological model. For example, only a single representative profile is extracted from simple flat-layer models, whereas multiple profiles are sampled from structurally complex models (e.g., Flat-Fault or Fold-Fault) to capture spatial variations. This strategy avoids over-representing simple structures while preserving the diversity of more complex geological settings.

#Modifications 1#:

2.3 OpenSWI-deep: Global Coverage Benchmark for Deep Earth Imaging

…, A quantitative analysis of the diversity and similarity of the extracted 1-D velocity profiles is provided in Appendix D.

Appendix D. Statistical Analysis of the Diversity of Extracted 1-D Velocity Models

To evaluate the structural diversity of the extracted 1-D velocity models and to assess potential similarity introduced by sampling multiple profiles from the same 3-D geological models, we conducted several statistical analyses on the velocity structures.

First, we evaluated the similarity between randomly sampled pairs of velocity profiles. A total of 100, 000 profile pairs were randomly selected from the extracted model library, and their differences were quantified using the L2 distance between shear-wave velocity vectors. Each profile was represented by a depth-sampled Vs vector with consistent sampling intervals. The resulting distribution of L2 distances (Fig. D1a) spans a broad range, indicating substantial structural variability among the velocity profiles despite being derived from a limited number of underlying 3-D models. This variability arises from both regional structural differences among the source geological models and the perturbation strategy applied during dataset augmentation.

Second, we performed a dimensionality reduction analysis using Principal Component Analysis (PCA) to visualize the global distribution of velocity structures. Each 1-D velocity profile was represented as a vector of shear-wave velocities sampled along depth and then projected into a two-dimensional principal component space. The PCA

distributions for velocity profiles derived from different source models are shown in Fig.D1 b. The PCA projections demonstrate that profiles originating from different regional models occupy distinct regions in the reduced feature space, reflecting systematic variations in crustal and upper mantle structures across different tectonic settings. Meanwhile, profiles extracted from the same regional model still exhibit a relatively broad spread in the PCA space, indicating that the perturbation strategy introduces additional structural variability while preserving the large-scale geological characteristics of the original models.

Overall, these statistical analyses suggest that the extracted and augmented 1-D velocity models cover a wide range of structurally diverse velocity profiles while maintaining geologically realistic constraints inherited from the underlying 3-D models. This balance between geological realism and structural variability is essential for constructing a robust benchmark dataset for surface-wave dispersion inversion.

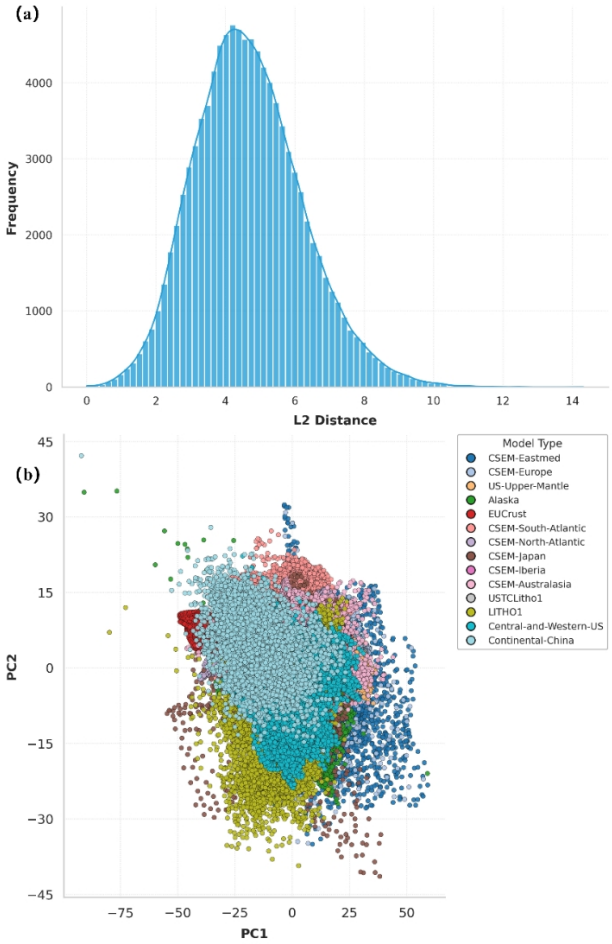

Figure D1. Statistical analysis of the structural diversity of the extracted 1-D velocity models. (a) Distribution of L2 distances between randomly sampled pairs of shear-wave velocity profiles (10^5 pairs), showing a broad range of structural differences among the extracted models. (b) PCA projection of the velocity profiles in

a two-dimensional feature space. Colors denote profiles derived from different source 3-D geological models, illustrating both the separation between regional structural patterns and the variability introduced by the perturbation strategy within each model group.

#Comment 2#: Too few information are provided, even in the appendix, about the DDPM. In particular, on how viable is to expand the dataset with diffusion model: does the DDPM reproduce the same statistics? how many iterations are needed to infer new samples? how diverse are those samples? Unless the DDPM model has some novel feature, I think its role in this paper is rather marginal and can be overlooked. Otherwise, it should be expanded to highlight its importance.

#Response 2#: We thank the reviewer for this helpful comment. We agree that the DDPM component is not the central contribution of this study, and its primary role is to provide an optional pathway for scalable dataset augmentation rather than to introduce a new generative modeling methodology. Our intention in including this module is to demonstrate how the OpenSWI-shallow dataset can be extended when additional geological variability is required.

Following the reviewer's suggestion, we have clarified the role of the DDPM and added additional information in the revised manuscript and appendix. First, we emphasize that the diffusion model is used only as an **optional dataset expansion tool**. The core OpenSWI-shallow dataset is constructed directly from the OpenFWI geological models, while the DDPM module provides an additional mechanism to generate structurally coherent velocity models when a larger number of samples is needed.

Second, we have expanded the appendix to further describe the statistical characteristics and diversity of the generated models. The DDPM is trained on the OpenFWI velocity models and therefore learns the statistical characteristics of the training dataset. As a result, the generated models reproduce the large-scale structural features present in the training data, such as layered sedimentary units, folds, and fault-related discontinuities, while introducing additional structural variations through the stochastic diffusion process. Specifically, we compared the distributions of the DDPM-generated velocity models and the original OpenFWI models using Principal Component Analysis (PCA) (Figure. R2). The results show that most generated models occupy similar regions of the PCA feature space as the training data, indicating that the DDPM successfully reproduces the overall structural statistics of the original dataset. At the same time, a portion of the generated samples extends beyond the main clusters of the training data, suggesting that the diffusion process also introduces additional structural variability and produces some out-of-distribution

samples. This property is beneficial for dataset expansion, as it increases the diversity of geological structures while maintaining consistency with realistic geological patterns. Furthermore, we have clarified the sampling procedure and computational cost. In the current implementation, velocity models are generated through a standard 1000-step denoising diffusion process. In practice, generating one 2-D velocity model requires approximately 0.35 seconds on a single Ascend 910B2 NPU, making it feasible to rapidly generate additional training samples when needed.

Finally, following the reviewer's suggestion, we revised the manuscript to explicitly describe the DDPM module as an optional dataset expansion strategy, rather than a core methodological contribution. This clarification helps place the diffusion model in the appropriate context within the dataset construction workflow.

#Modifications 2#:

2.2.2 Optional Dataset Expansion with DDPM

Although the proposed OpenSWI-shallow dataset constructed from OpenFWI substantially improves geological structural diversity compared with existing dispersion curve datasets, it cannot fully cover the complete range of velocity structure observed in real subsurface settings. To provide a scalable pathway for further dataset expansion, we optionally incorporated a deep generative module based on Diffusion Probabilistic Model (DDPM), specifically designed for the shallow subsurface within the 0-3 km depth range.

C4. DDPM sampling and OpenSWI-shallow datasets Generation

…, In practice, generating one 2D velocity model requires approximately 0.35 seconds on a single Ascend 910B2 NPU, making it feasible to rapidly expand the dataset when needed.

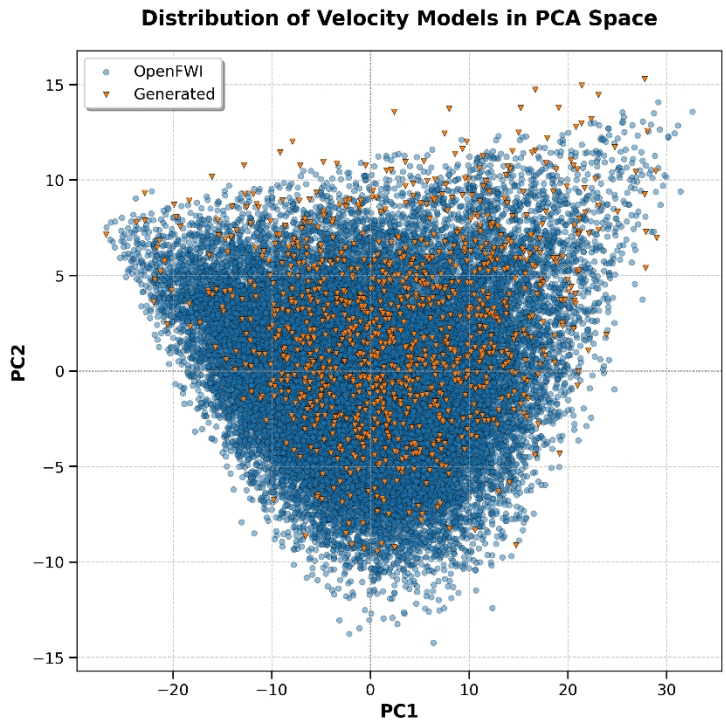

Figure R2. Statistical comparison between OpenFWI velocity models and DDPM-generated velocity models for the curve-vel dataset. The figure shows the PCA projection of velocity models in the reduced feature space. Blue dots represent the original OpenFWI velocity models, and orange triangles denote the DDPM-generated models. The strong overlap between the two distributions indicates that the generated models reproduce the statistical characteristics of the training dataset while introducing additional structural variability.

#Comment 3#: what is the highest frequency that the geological models can propagate?

#Response 3#: We thank the reviewer for raising this important point regarding the frequency limits supported by the geological models in OpenSWI. The physically meaningful frequency range for surface-wave dispersion is primarily determined by the vertical resolution of the velocity models, i.e., the minimum layer thickness $h_{\min}$, and the local shear-wave velocity V_s^{min}. We estimated the theoretical maximum frequency for each model using

$$f_{max} \approx V_s^{min}/(2\,h_{min}).$$

As shown in Figure R3, for **OpenSWI-shallow**, with a minimum layer thickness of 40 m, the resulting $f_{\max}$ distribution spans approximately 3–20 Hz, with most models concentrated around 5 Hz. For **OpenSWI-deep**, with a minimum layer

thickness of 1 km, the $f_{\max}$distribution ranges from 0.5–2.0 Hz, peaking near 1 Hz.

In comparison, the dispersion curves provided in the dataset are generated for periods of 0.2–10 s (corresponding to 0.1–5 Hz) for OpenSWI-shallow, and 1–100 s (0.01–1 Hz) for OpenSWI-deep. These selected period ranges are below the theoretical maximum frequencies of the models, ensuring that all generated surface-wave dispersion curves are physically meaningful and consistent with the structural resolution of the underlying velocity models.

#Modifications 3#:

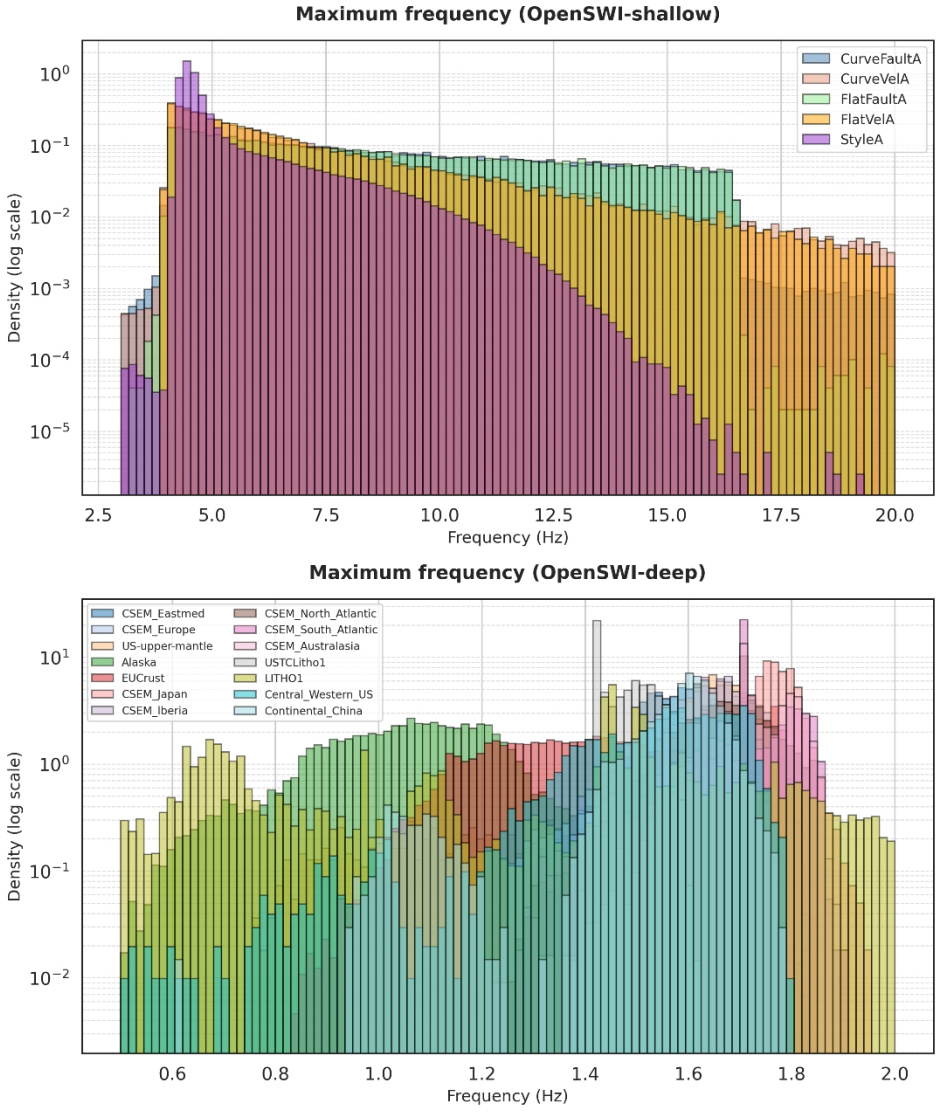

Figure R3. Theoretical maximum frequency distribution of velocity models in OpenSWI. The distributions of $f_{\max} \approx V_s^{\min}/(2h_{\min})$ are shown for OpenSWI-shallow (top) and OpenSWI-deep (bottom).

#Comment 4#: Are the random perturbations introduced by author consistent with the natural uncertainty? What about small scale heterogeneity which is well known to have a specific 3D correlation structure? Why did not the authors include this in their dataset?

#Response 4#: We thank the reviewer for raising this insightful question regarding the realism of the perturbation strategy and the treatment of small-scale heterogeneity. The perturbations introduced in the OpenSWI dataset are designed to represent the natural uncertainty commonly present in subsurface seismic velocity structures, consistent with strategies adopted in previous studies (Luo et al., 2022; Huang et al., 2024; Liu et al., 2025). Their amplitudes are defined within geologically reasonable ranges based on typical variability observed in seismic velocity models.

The primary objective of the OpenSWI dataset is to support surface-wave dispersion inversion, which mainly targets the recovery of 1-D shear-wave velocity (Vs) profiles. Accordingly, the dataset is constructed based on 1-D velocity models extracted from geological structures. As a unified modeling strategy, we introduce controlled perturbations directly to these extracted 1-D velocity profiles. The perturbation ranges are defined relative to the original velocity models, ensuring that the large-scale geological structure remains physically reasonable while allowing local variations to increase the diversity of velocity profiles.

The reviewer also points out that small-scale heterogeneity in the Earth often exhibits spatially correlated three-dimensional structures. We agree that such heterogeneity can play an important role in wave scattering and high-frequency waveform simulations. In the context of the OpenSWI dataset, which is designed specifically for 1-D surface-wave dispersion inversion, representing spatially correlated 3-D heterogeneity is beyond the intended scope. Incorporating such structures would require a full three-dimensional model representation and would substantially increase the complexity of the dataset without directly contributing to the target 1-D velocity inversion problem. For this reason, the current dataset focuses on layered velocity structures with controlled perturbations applied to the extracted 1-D profiles. This approach preserves the geological realism of the underlying models while introducing sufficient structural variability for training and benchmarking dispersion inversion methods. Incorporating spatially correlated heterogeneity could be valuable for future studies focusing on waveform modeling or scattering effects, and we consider this a potential direction for future extensions of the dataset.

#Modifications 4#

2.1.3 Augmentation of Velocity Models for Geological Diversity

…, applying constrained perturbations to the velocity and thickness of each layer

within predefined ranges to generate structurally consistent variations (Luo et al., 2022; Huang et al., 2024; Liu et al., 2025).

Reference:

[1] Luo, Y., Huang, Y., Yang, Y., Zhao, K., Yang, X., and Xu, H.: Constructing Shear Velocity Models from Surface Wave Dispersion Curves Using Deep Learning, Journal of Applied Geophysics, 196, 104 524, https://doi.org/10.1016/j.jappgeo.2021.104524, 2022.

[2] Huang, X., Yu, Z., Wang, W., and Wang, F.: JointNet: A Multimodal Deep Learning-Based Approach for Joint Inversion of Rayleigh Wave Dispersion and Ellipticity, Bulletin of the Seismological Society of America, 114, 627 – 641, https://doi.org/10.1785/0120230199, 2024.

[3] Liu, F., Deng, B., Su, R., Bai, L., and Ouyang, W.: DispFormer: Pretrained Transformer for Flexible Dispersion Curve Inversion from Global Synthesis to Regional Applications, https://doi.org/10.48550/ARXIV.2501.04366, 2025.

#Comment 5#: The authors overlooked one major dataset, published on this journal in 2024, which provides 30000 ground motion simulations including complex randomized geology: Lehmann, F.; Gatti, F.; Bertin, M.; Clouteau, D. Synthetic Ground Motions in Heterogeneous Geologies from Various Sources: The HEMEW S -3D Database. Earth Syst. Sci. Data 2024, 16 (9), 3949 3972. https://doi.org/10.5194/essd-16-3949-2024. This database span a ~10x10 km² for each sample and it is constructed with a minimum bias. Considered the fact that the dataset provides (geology, time-histories) couples, it would be interesting to benchmark the proposed model out-of-distribution, which is the most difficult aspect of benchmarking a new ML model

#Response 5#: We thank the reviewer for highlighting the HEMEW-S-3D database (Lehmann et al., 2024) and for suggesting the possibility of evaluating the proposed framework on such datasets. The HEMEW-S-3D database is indeed a valuable resource that provides large-scale ground-motion simulations in complex heterogeneous geological environments and represents an important contribution to the community.

The OpenSWI-shallow dataset focuses on near-surface shear-wave velocity structures within a depth range of approximately 0-2.8 km. In contrast, the HEMEW-S-3D database contains three-dimensional heterogeneous geological models spanning areas of approximately 10 × 10 km² for each simulation and provides corresponding ground-motion time histories. This difference in spatial scale and data representation means that the two datasets target somewhat different levels of

geological description. Directly applying the current OpenSWI workflow to HEMEW-S-3D would therefore require additional processing steps, such as extracting surface-wave components from the simulated wavefields and estimating dispersion curves from the resulting time series.

Nevertheless, we agree that datasets such as HEMEW-S-3D provide valuable opportunities for evaluating the generalization capability of data-driven inversion approaches under more complex geological conditions. In particular, the heterogeneous structures represented in HEMEW-S-3D could serve as an important resource for future out-of-distribution benchmarking and for further expanding the geological diversity of dispersion datasets.

#Modifications 5#

1. Introduction

Seismic exploration and engineering have also benefited from the development of standardized workflows and open benchmark datasets, such as cigFacies (Gao et al., 2025), cigChannels (Wang et al., 2025), and the HEMEWS-3D database for large-scale ground motion simulations in heterogeneous geological environments (Lehmann et al., 2024).

4. Discussion

Future developments can be pursued along several interrelated directions. First, expanding the dataset's geographic coverage and geological diversity, particularly in tectonically extreme regions, would broaden its applicability. In particular, large-scale synthetic datasets incorporating heterogeneous three-dimensional geological structures, such as the HEMEWS-3D database (Lehmann et al., 2024), provide valuable resources for constructing more complex training and benchmarking scenarios.

Reference:

Lehmann, F., Gatti, F., Bertin, M., and Clouteau, D.: Synthetic Ground Motions in Heterogeneous Geologies from Various Sources: The HEMEWS-3D Database, Earth System Science Data, 16, 3949 – 3972, https://doi.org/10.5194/essd-16-3949-2024, 2024.

#Comment 6#: The transformer architecture presented in the paper seem a little too advanced for such a simple dataset (dispersion curves vs 1D geological profile). It is necessary to benchmark it with existing alternative deep learning

models in order to consider it as a reliable alternative.

#Response 5#: Thank you for this important comment. Benchmarking against alternative neural network architectures is indeed necessary when introducing a transformer-based approach for dispersion-curve inversion. In response to this suggestion, additional benchmarking experiments have been conducted in the revised manuscript. Specifically, two representative deep learning architectures that have previously been applied to surface-wave inversion were implemented for comparison: a U-Net-based convolutional neural network and a fully connected neural network (FCNN). These models were trained and evaluated on the OpenSWI-shallow and OpenSWI-deep datasets under identical training configurations. The benchmarking results are presented in the Appendix of the revised manuscript. The comparison shows that while the CNN/U-Net and FCNN models achieve reasonable performance on individual datasets, the transformer-based architecture consistently yields lower RMSE values. Beyond the quantitative accuracy, an important practical distinction lies in the ability of the models to handle dispersion curves with variable sampling characteristics. Conventional CNN and FCNN architectures require fixed-length input representations and therefore cannot be directly applied to dispersion curves with varying period ranges or sampling densities, such as those encountered in the OpenSWI-real dataset. In contrast, the transformer architecture naturally supports variable-length input sequences and can therefore be applied directly to observational dispersion curves without additional preprocessing or retraining. The use of a transformer-based architecture is also consistent with several recent studies that have successfully applied attention-based models to dispersion-curve inversion problems (Huang et al., 2024; Liu et al., 2025; Jiang et al., 2025). These works suggest that attention mechanisms are well suited for capturing the complex relationships between dispersion curves and subsurface velocity structures. The additional benchmarking results included in the revised manuscript therefore provide a clearer comparison with existing architectures and further support the suitability of the transformer-based approach for this task.

#Modifications 6#

Appendix F: Benchmarking Alternative Neural Network Architectures

F1. Compared Architectures

To assess the effectiveness of the proposed approach, three representative neural network architectures previously applied to surface-wave dispersion curve inversion are considered: a U-Net-based model (Wang et al., 2023b), a fully connected neural network (FCNN) (Chen et al., 2024), and the Transformer-based architecture adopted

in this study. The U-Net architecture is a convolutional encoder – decoder network originally developed for image segmentation and subsequently adapted to geophysical inversion problems. In this benchmark, a one-dimensional U-Net implementation following the design proposed by Wang et al., (2023b) is adopted. The model consists of four encoder – decoder stages with skip connections, where convolutional layers progressively extract hierarchical features from the dispersion curves and reconstruct the corresponding subsurface shear-wave velocity profiles. The FCNN model follows the architecture described by Chen et al., (2024). It consists of an initial convolutional layer serving as a feature embedding module, followed by seven fully connected layers that map dispersion-curve features directly to the target shear-wave velocity profile. Detailed architectural configurations of the U-Net and FCNN models are available in the corresponding references. In the present benchmark, both models are implemented following the configurations described in the original studies to maintain consistency with previous work.

F2. Experimental Setup

The CNN/U-Net and FCNN architectures require fixed-length input representations. As a result, these models cannot be directly applied to dispersion curves with variable sampling densities or period ranges, such as those present in the OpenSWI-real dataset. Consequently, the benchmarking experiments are conducted exclusively on the OpenSWI-shallow and OpenSWI-deep datasets. To ensure a fair comparison across different architectures, several training strategies employed in the main experiments are intentionally simplified. In particular, no additional data augmentation techniques are applied in the benchmarking experiments, including the depth-aware masking strategy and the random noise injection described in the main text.

All models are trained using an identical dataset partitioning strategy, consisting of 90%, 5%, and 5% splits for training, validation, and testing, respectively. The evaluation results reported here correspond to the performance on the held-out 5% test subset. To further ensure consistency, identical optimization settings are adopted for all models. Specifically, the Adam optimizer is used with an initial learning rate of 0.0001, combined with a warm-up phase followed by a step-based learning rate decay schedule (StepLR). The maximum number of training epochs is set to 30 for the OpenSWI-shallow dataset and 200 for the OpenSWI-deep dataset. To examine the potential influence of the training objective, two commonly used regression loss functions are considered: mean squared error (MSE) and mean absolute error (MAE). Each network architecture is trained separately using both loss functions under identical training configurations, resulting in six benchmarking experiments (three

network architectures combined with two loss functions). For consistency, the evaluation metric reported in the comparison is the root mean square error (RMSE) between the predicted and reference shear-wave velocity profiles on the test dataset.

F3. Results and Discussion

Table F1 summarizes the benchmarking results obtained using different network architectures and loss functions on the OpenSWI datasets. The results indicate that the Transformer-based architecture consistently achieves the lowest RMSE across both datasets and loss-function settings. The U-Net model exhibits comparable performance on the OpenSWI-shallow dataset but shows larger errors on the more challenging OpenSWI-deep dataset. In contrast, the FCNN model yields relatively higher errors overall, suggesting that its limited representational capacity may restrict its ability to capture the complex nonlinear relationships between dispersion curves and subsurface velocity structures. Regarding the influence of the loss function, the RMSE values obtained using MSE and MAE are generally similar, with only minor variations between the two settings. This observation suggests that the overall inversion performance is primarily governed by the network architecture rather than the specific regression loss used during training.

Beyond the quantitative accuracy presented in Table F1, an important practical distinction lies in the ability of different architectures to generalize to real observational datasets. The CNN/U-Net and FCNN models require fixed-length input representations and therefore cannot be directly applied to dispersion curves with varying sampling densities or period ranges, such as those encountered in the OpenSWI-real dataset. In contrast, the Transformer-based architecture naturally supports variable-length input sequences and can therefore be applied directly to real observational dispersion curves without additional preprocessing or retraining. These results highlight an important consideration for future deep-learning-based surface-wave inversion methods: in addition to achieving strong performance on synthetic benchmark datasets, inversion models should also possess sufficient flexibility to accommodate dispersion curves with varying period ranges and sampling densities commonly encountered in real-world applications.

Table F1. Benchmark comparison of different neural network architectures and loss functions on the OpenSWI datasets. Values represent RMSE (km/s) computed on the held-out test subsets.

Dataset	U-Net (MAE)	U-Net (MSE)	FCNN (MAE)	FCNN (MSE)	Transformer (MAE)	Transformer (MSE)
OpenSWI-shallow	0.1407	0.1413	0.2366	0.2269	0.1411	**0.1353**
OpenSWI-deep	0.0454	0.0421	0.0617	0.0554	0.0164	**0.0163**